

Serial No. 10/059,421
Docket No. YOR920010126US2

2

RECEIVED
CENTRAL FAX CENTER

FEB 19 2008

AMENDMENTS TO THE CLAIMS:

Please cancel claims 12 and 24 without prejudice or disclaimer.

1-30. (Canceled)

31. (New) A system for identifying genes, comprising:
a pattern database comprising patterns of amino acids;
an input device for inputting a genomic DNA sequence; and
a processor which:
translates an open reading frame (ORF) of said DNA sequence into an amino acid translation;
locates in said amino acid translation occurrences of said patterns from said pattern database; and
determines whether said open reading frame includes a putative gene based on a number of said patterns of amino acids located in said amino acid translation of said ORF, and/or weighted values associated with said patterns of amino acids located in said amino acid translation of said ORF.
32. (New) The system according to claim 31, wherein said processor translates a plurality of open reading frames in said DNA sequence into amino acid translations, and locates in each amino acid translation occurrences of said patterns to determine whether each said plurality open reading frames includes a putative gene.
33. (New) The system according to claim 32, wherein said patterns comprise biologically significant patterns of amino acids in amino acid sequences.
34. (New) The system according to claim 31, wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation.

Serial No. 10/059,421
Docket No. YOR920010126US2

3

35. (New) The system according to claim 34, wherein said patterns are derived from a parent database comprising at least one amino acid sequence.

36. (New) The system according to claim 34, wherein said patterns are derived from a parent database comprising at least one amino acid sequence fragment.

37. (New) The system according to claim 34, wherein said patterns are derived by using a pattern discovery algorithm.

38. (New) The system according to claim 34, wherein said patterns are derived by using the Teiresias algorithm.

39. (New) The system according to claim 34, wherein said ORF comprises a portion of said DNA sequence between a start codon and a stop codon.

40. (New) The system according to claim 34, wherein said processor reports said ORF as a putative gene when a predetermined number of pattern matches is identified in said amino acid translation.

41. (New) The system according to claim 34, wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene.

42. (New) The system according to claim 34, wherein said ORF is reported as a putative gene when the sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.

Serial No. 10/059,421

4

Docket No. YOR920010126US2

43. (New) The system according to claim 34, wherein said match is identified using a predetermined pattern matching algorithm.
44. (New) The system according to claim 34, further comprising:
a memory device for storing data and instructions to be executed by said processor.
45. (New) The system according to claim 34, further comprising:
a display device for displaying an output from said processor.
46. (New) A method of identifying genes, comprising:
providing a pattern database comprising patterns of amino acids;
determining an open reading frame (ORF) in a genomic DNA sequence;
generating an amino acid translation for said ORF;
locating a match of a pattern from said pattern database in said amino acid translation;
determining whether said ORF includes a putative gene based on a number of said matching patterns of amino acids located in said amino acid translation of said ORF, and/or a weighted value associated with a matching pattern located in said amino acid translation of said ORF; and
displaying a result of said determining whether said ORF includes a putative gene.
47. (New) The method according to claim 46, wherein said pattern database is generated from a database comprising at least one amino acid sequence.
48. (New) The method according to claim 46, wherein said pattern database is generated from a database comprising at least one amino acid sequence fragment.
49. (New) The method according to claim 46, further comprising:
identifying said ORF as a putative gene when a predetermined number of pattern

Serial No. 10/059,421

5

Docket No. YOR920010126US2

matches is identified in said amino acid translation.

50. (New) The method according to claim 46, further comprising:
assigning a weight to each pattern depending upon a relevance of said pattern in
determining whether said ORF includes a putative gene.
51. (New) The method according to claim 46, further comprising:
displaying said match of said pattern in said amino acid translation.
52. (New) The method according to claim 46, wherein said pattern database is generated
using the Teiresias algorithm to derive said patterns from a parent database.
53. (New) A programmable storage medium tangibly embodying a program of
machine-readable instructions executable by a digital processing apparatus to perform a method
for identifying genes, said method comprising:
providing a pattern database comprising patterns of amino acids;
determining an open reading frames (ORF) in a given genomic DNA sequence;
generating an amino acid translation for each ORF;
locating a match of a pattern from said pattern database in said amino acid
translation;
determining whether said ORF includes a putative gene based on a number of
matching patterns of amino acids located in said amino acid translation of said ORF, and/or a
weighted value associated with a matching pattern located in said amino acid translation of said
ORF; and
displaying a result of said determining whether said ORF includes a putative
gene.
54. (New) The system according to claim 33, wherein said processor

Serial No. 10/059,421
Docket No. YOR920010126US2

6

determines for each pattern in said pattern database whether the pattern is present in said amino acid translation by locating instances of said patterns in said amino acid translation, until a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold, at which point said processor identifies said ORF as a putative gene.

55. (New) The system according to claim 31, further comprising:

a parent database comprising a plurality of amino acid sequences, said patterns in said pattern database being derived from said plurality of amino acid sequences by using a pattern discovery algorithm;

a memory device for storing data and instructions to be executed by said processor; and

a display device for displaying an output from said processor.

56. (New) The system according to claim 55, wherein said open reading frame (ORF)

comprises a portion of said DNA sequence between a start codon and a stop codon,

wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation by using a predetermined pattern matching algorithm,

wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene, and

wherein said ORF is reported as a putative gene when either a predetermined number of pattern matches is identified in said amino acid translation, or a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.

57. (New) The system according to claim 31, wherein said processor accesses said pattern database to retrieve said patterns from said pattern database.

58. (New) The system according to claim 31, wherein said processor is electrically coupled to said input device and said pattern database.

Serial No. 10/059,421

7

Docket No. YOR920010126US2

59. (New) A system for identifying genes, comprising:
an input device which inputs a genomic DNA sequence; and
a processor which:
accesses a pattern database comprising a plurality of patterns of amino acids;
translates an open reading frame (ORF) of said DNA sequence into an amino acid translation;
locates in said amino acid translation occurrences of said patterns from said pattern database; and
determines whether said open reading frame includes a putative gene based on a number of said patterns of amino acids located in said amino acid translation of said ORF, and/or weighted values associated with said patterns of amino acids located in said amino acid translation of said ORF.
60. (New) A system for identifying genes, comprising:
an input device which inputs a query genomic DNA sequence;
a processor which:
accesses a pattern database comprising a plurality of patterns of amino acids;
translates an open reading frame (ORF) of said DNA sequence into an amino acid translation; and
locates in said amino acid translation occurrences of said patterns from said pattern database;
determines whether said open reading frame includes a putative gene based on a number of said patterns of amino acids located in said amino acid translation of said ORF, and/or weighted values associated with said patterns of amino acids located in said amino acid translation of said ORF; and
a display device for displaying an output of said processor, said output including an occurrence of said patterns in said amino acid translation,
wherein said patterns comprises patterns derived using a Teiresias algorithm,

Serial No. 10/059,421

8

Docket No. YOR920010126US2

wherein said open reading frame (ORF) comprises a portion of said DNA sequence between a start codon and a stop codon,

wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation by using a predetermined pattern matching algorithm,

wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene, and

wherein said ORF is reported as a putative gene when either a predetermined number of pattern matches is identified in said amino acid translation, or a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.